

Importance of Databases in Experimental and Clinical Allergology

Adriano Mari

Allergy Data Laboratories s.c., Latina, Italy

Key Words

Allergen database · Internet · Sequence homology collection · Information technology · Proteomic microarrays

Abstract

Information technology (IT) is leading us to reconsider some of the approaches we have been using in both basic research and clinical work in allergology. Resources mainly coming from the advent of the Internet are further amplified by the parallel development of other new tools, such as molecular biology and nanotechnology. These three powerful tools are now available and are cross-linked to a certain degree to express their power when applied to biomedical fields. Bioinformatics applied to allergy simplifies our way of handling an increasing wealth of knowledge. This review assesses the current status of allergen databases that are mainly dedicated to sequence homology collection for computational purposes. Whether or not they integrate features that are now typical of IT in other biomedical fields is analyzed as well. The results of these analyses are discussed with a view to the critical need of integrating biochemical data with clinical, epidemiological information and how this goal can be reached by the use of proteomic microarrays for IgE detection. Future directions for a more comprehensive use of allergen databases are proposed.

Copyright © 2005 S. Karger AG, Basel

Introduction

In the last 20 years three new tools have started to be applied in biomedical research and in clinical practice: molecular biology, nanotechnology, and information technology (IT). Allergology, with its peculiarity of being linked to environmental organism products causing allergic diseases, is on the way to define its own applications of these new tools. Molecular biology and all the biochemical and immunochemical techniques have enabled the identification of a number of structures leading to increasing knowledge on allergens [1–4]. The time trend of these historical changes, summarized by Harwanegg et al. [5], is also depicted in figure 1, where the cumulative number of allergens identified during the last 40 years is reported. The application of nanotechnology in allergology is in its early infancy. In relation to IgE sensitizing/triggering compounds, it is also used in the creation of proteomic microarrays. They have been based on classical allergenic extracts [6–11], but more recently native and recombinant purified allergens have been successfully tested [5, 12–16] and are going to definitely replace extracts. The use of microarrays in allergy is expected to change the approach to epidemiology and diagnosis of IgE-mediated diseases [5]. Nanotechnology is essentially a high-throughput technology, and the products of the combined use of allergenic molecules produced by means of molecular biology techniques and microarrays require management and interpretation. In parallel, IT, the third new tool

mainly supported by the Internet, has allowed us to create online databases that can be easily accessed worldwide. In the last two decades, several research groups and institutions have started to collect information and data from available sources in order to create databases of scientific knowledge. For instance, a growing list of molecular biology databases is reported every year [17]. A generic database is defined as being capable of acquiring, storing, visualizing, analyzing and interpreting data (fig. 2). These steps are applicable to information and data processing in any biomedical field. The present paper will review the state of the art of allergy databases available

on the Internet, and how they integrate in their structure one or more aspects of a generic database. The review will first focus on allergen sequence databases as they are the more numerous, and then explore the potential application of IT in allergology – mainly literature mining and data mining (fig. 2). Web sites considered in the present review deal with protein antigens recognized by human IgE. Only free access noncommercial resources have been reviewed. Table 1 reports comparative features of these online resources as they are on the web or described in scientific journals [18–22].

Allergen Lists, Sequence Databases and Computational Tools

Information on allergenic sources, namely organisms inducing IgE-mediated diseases, historically date back to the beginning of the last century. Compared with those available for allergen molecules they are far more extensive. Nevertheless, historically comprehensive lists or databases of allergenic sources are not available online yet. The Allergome database for all allergens, and the Inform All database for food allergens are trying to define a list of organisms inducing allergies on the basis of the published literature. Although it may seem outdated to classify organisms capable of inducing allergic reactions rather than their molecular structures, the still expanding knowledge on allergenic sources is a continuous challenge for achieving the most comprehensive list of allergen molecules. It is also important to keep in mind that the emergence of allergy as a problem in developing countries and the globalization in terms of mobility of people and goods

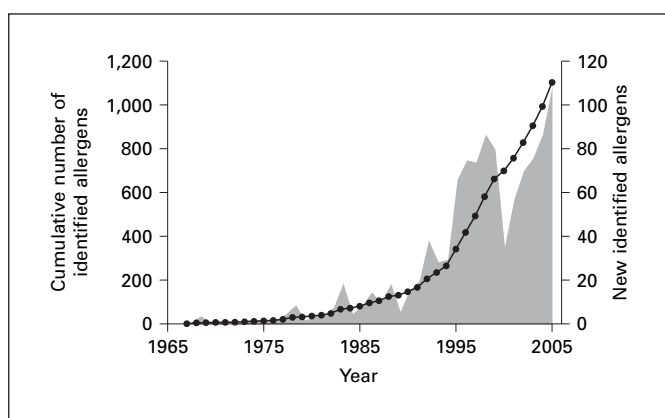


Fig. 1. Identification of new allergens during the last 40 years. The black line shows the cumulative number of allergens (left Y axis). The grey area is the year by year number of identified allergens (right Y axis). Data are extracted from the Allergome database (www.allergome.org).

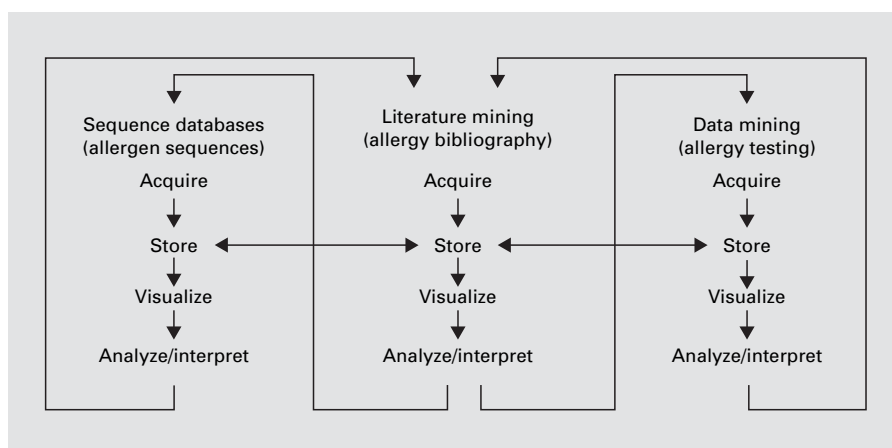


Fig. 2. Main fields and future interests in allergy database development, and their interrelated feeding.

create new allergenic sources [23–25] and new routes of exposure [26]. Thus, a reliable database resource for allergenic organisms should be established as well. For each organism, the molecular perspective should enable us to define whether we are dealing with either a new allergen [27] or with a different source containing a formerly known allergenic structure [28]. This classification process based on molecular allergens would also lead to reconsider several well-known allergenic organisms (i.e. grass species) as part of a unique cluster of immune reactivity, having IgE corecognized structures [4, 29].

The first list of allergens available from the web was released by the IUIS subcommittee for allergen nomenclature in 1996, as a downloadable file from an ftp server

[30]. Later on, this list was transformed into the IUIS official nomenclature web site in which submitted allergen sequences are named according to official criteria first defined in 1986 [31], and revised in 1995 [2, 32]. The official list is currently updated two to four times per year and contains elements for allergen and author identification. The activities of the IUIS subcommittee and their web product play a crucial role for the scientific community, as they should avoid arbitrary assignment of the same allergen name by different authors. The same work is carried out for similar aspects within the immune systems, for the associated CD nomenclature that is commonly used when referring to leukocyte surface molecules and antibodies against them. It provides an essential clas-

Table 1. Main features of Internet-available resources on allergens

Web site	URL	Data Source	Allergen sources	Allergen molecules	Allergen sequence database	Computational tools	Bio-chemistry molecular biology	Clinical/epidemiological data	Internal search engine	Internal data cross-linking	User alert	Ref.	Last update
Official allergen nomenclature of WHO-IUIS ¹	www.allergen.org	submission	no	all	no	no	+	no	no	no	no	+	Jun 02, 2005
Allergen Database (ADB)	http://allergen.csl.gov.uk/index.htm	sequence databases	no	all	external links	no	+	no	yes	no	no	+	given for each allergen
Allergen Online (FARRP)	www.allergenonline.com	sequence databases	no	all	yes & ext. links	yes & external	++	no	yes	no	yes	no	Oct 1, 2004
Allergome	www.allergome.org/	literature & sequence databases	all	all	external links	no	+++	yes	yes	yes	yes	+++	Jul 17, 2005 & given for each allergen
AllerMatch	www.allermatch.org/	sequence databases	no	all	yes & ext. links	yes	++	no	no	no	no	no	Dec 10, 2004
AllerPredict	http://sdmc.i2r.a-star.edu.sg/Templar/DB/Allergen/	sequence databases	no	all	yes & ext. links	yes & external	+	no	yes	yes	no	+	given for each allergen
Bioinformatics for Food Safety (BIFS)	http://www.iit.edu/~sgendel/fa.htm	sequence databases	no	all	external links	no	no	no	no	no	no	+	not given
InformAll (Protall)	http://foodallergens.ifrac.ac.uk/	literature & sequence databases	food	food	external links	no	+++	yes	yes	no	no	+++	given for each allergen
Structural Database of Allergen Proteins (SDAP)	http://fermi.utmb.edu/SDAP/sdap_src.html	sequence databases	no	all	yes & ext. links	yes & external	+++	no	yes	no	yes	no	given for each allergen
WebAllergen	http://weballergen.bii.a-star.edu.sg/	sequence databases	no	all	yes	yes	no	no	no	no	no	no	not given

+, ++, +++ = Arbitrary grading of the supported computational tools, and other supplied information.

¹ Official web site of the WHO/IUIS allergen nomenclature subcommittee.

sification for diagnostic and therapeutic purposes [33]. For the same reason, the IUIS web site for allergen nomenclature should be considered essential by the international scientific community. It is expected that the main limitation of the IUIS list due to incomplete listing of described allergens, mainly due to nonsubmission by authors, could be overcome by defining different strategies of data collection for the future. This should also include the definition of a molecule as an allergen by updating the current criteria [32] and including several new parameters. It would enable to grade the allergenicity of novel and established allergens. The dynamic nature of IT could help in this respect.

The Bioinformatics for Food Safety database (BIFS), released in 1998, was the first to attempt to list allergens and create a database for computational purposes [18]. At the moment, the BIFS web site lists food and nonfood allergens whose sequences are available from linked external databases. Many allergen isoforms and variants are listed, but, if compared with the IUIS web list (as of June 2005), it seems not to be updated in terms of both listing of unique allergens and their nomenclature. The last update is not given. Furthermore, no original or shared computational tools are provided within the BIFS web site.

The Allergen Database (ADB) has been set up in 2000 and contains information regarding allergen identification. Links to protein sequence databases are available in allergen monographs. For some allergens they include epitope sequences. Compared with the last release of the IUIS web list, the ADB seems not to be updated in terms of listed allergens and their nomenclature. The update is given for each available allergen. No computational tools are available.

The PROTALL database was created in 2001 in the frame of a European Union funded project (FAIR-CT98-4356). The PROTALL database is now the basis of a new database, the InformAll, another European Union project funded under the 5th Framework programme through the Quality of Life and Management of Living Resources Key Action. The aim of this project is to create a credible food allergy information source to be used by consumers, regulators and industry. The first release of the database was in January 2005 and the final release is expected to be completed by the end of 2005. Inclusion criteria and explanations on monograph sections are carefully detailed on the web site. The InformAll database is based on allergenic source monographs, which allow identification of the source. Pages dedicated to the identification of allergenic molecules are linked in each source monograph. They report exhaustive biochemical and clinical

data, including characterized epitopes, in a text format resembling a peer reviewed paper. The InformAll focuses solely on food allergens with cross-citations to other non-food allergens within monographs. A friendly interface eases the access to data. The development of computational tools is not planned within the InformAll project.

The Structural Database of Allergen Proteins (SDAP), first released in 2001 [34], is one of the most up-to-date and complete lists of allergen sequences. SDAP also includes identified B epitopes from allergens. It includes all the IUIS allergens, which are marked, as well as allergenic structures whose sequences are available in other protein databases. Some IgE-binding structures from parasites are listed as well. The database supplies a battery of powerful computational tools available from other web sites (FASTA, BLAST) or originally released by SDAP developers. These computational tools allow comparison of entire allergen sequences or their peptides against the SDAP database or external protein databases [35]. The SDAP is the only allergen database listed in the annual database issue of *Nucleic Acids Research* [17].

The AllergenOnline is a database created in 2002 within the Food Allergy Research and Resource Programme. It collects sequences from food and nonfood allergens. Parasite IgE-binding proteins are listed as well. Non-IUIS allergens are included. The sequence database is released once a year. Computational tools available within this web site include a FASTA search for sequence homology and an original algorithm [36] was recently further implemented with an 80 mer search.

The Allergome database was released in 2003. It is the sole dynamic database among those considered in the present review. Data are stored in archives. The original software allows retrieval of the last updated information immediately after it is saved in one of the archives and linked to an allergen [20]. The aim of the Allergome database is to collect basic and clinical information as well as data about any allergenic organism and its molecular structures capable of inducing an IgE-mediated disease in a single web-based resource. Each allergen described in the dynamic monograph includes data to identify the allergen by means of its updated nomenclature, the source, and the biochemical and immunochemical features by means of a number of given or linked information. Each monograph is comprehensively referenced. All IUIS allergens are included and the Allergome database is updated as soon as a new release of the IUIS list is on the web. Other data sources to keep the database up to date are the scientific literature, which is scanned on a daily basis, and sequence databases. In spite of this, authors

report that papers published earlier than 1987 are still not completely processed. Non-IUIS allergens are included in the Allergome database as well. IUIS allergens are marked and searches can be performed excluding non-IUIS allergens. Allergome developers are attempting to provide the web site with a friendly interface to allow users to easily access information. A powerful internal search engine allows the user to navigate within the web site without launching a new search each time. No computational tools have been developed within the Allergome web site. A collaborative study has been established with researchers working in a Swiss immunology department who developed an original algorithm attempting to define the allergenic nature of molecules on the basis of motif recognition [37]. The motif discovery is now run every month by using updated lists of sequences from the Allergome database. The obtained motifs are then displayed in allergen monographs, and can be both searched and used for internal searches [38]. In an attempt to improve visualization of information in the database, an algorithm to generate a graph representing both sequence homology and IgE corecognition has been developed [39]. This tool, the O-ring, can be activated within each allergen monograph and dynamically scans the sequence homology and the IgE co-recognition sections to visualize allergen reciprocal relationships.

AllerPredict web site, released in 2003, is structured in cross-linked allergen monographs. Amino acid sequences are given for each molecule, as well as organism identification and external links. Allergens in this database do not seem to be fully updated compared with the IUIS web list. The update is given for each monograph, and unless not clear from the information page, it seems to work in a dynamic-like way. Two computational tools are available. An original algorithm enables comparison of a sequence with the AllerPredict software. A BLAST search is provided as a second tool for comparative searches.

Allermatch and WebAllergen sites, both released in 2004, are mainly dedicated to sequence comparison for allergenicity definition. Allermatch works on a database compiled on the basis of the allergen sequences available on Swissprot that is integrated with the IUIS list. The last update is given on the web site. WebAllergen is the last allergen database released on the web [22]. The compiled allergen database was designed to include a dataset of allergenic motifs. New sequences are compared with the motif set and with sequences lacking a motif recognition. The original software allows users to upload their own sequences. The software will produce a new set of motifs.

Comparison can also be performed against the new motif set.

Another computational resource not available on the web but released in the last few years has been published [40]. The Detection based on Automated Selection of Allergen-Representative Peptides (DASARP) is a further attempt to define the allergenicity of proteins based on a data set from other allergen databases and on a set of molecules considered by the authors to be devoid of allergenic activity [40]. The DASARP is not directly available on the web, but authors may be contacted for allergenicity assessment of specific protein sequences of interest. The full data set of allergenic proteins used in the study can be downloaded without restrictions.

Summarizing the current status and the trends in allergen sequence databases, it must be said that with few exceptions they have been mainly set up for computational purposes related to definition of allergenicity rather than to create general information resources for allergology.

Literature Mining

The publication of an increasing number of original papers in scientific journals is both the cause and the effect of our knowledge in biomedicine. The most used online resource for bibliography documentation, PubMed, from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) now hosts more than sixteen million references from worldwide biomedical journals published during the last 50 years. There is a critical need for extraction, integration and management of information discovery and synthesis from these sometimes highly heterogeneous data [41]. There is now a specific research field dedicated to developing informatics tools for literature mining [42]. These tools should allow us to summarize major themes and topics, evaluate large result sets quickly, and reveal trends and association in references. The same time trend of biomedical publications is also observed in allergology. Figure 3 reports the distribution of studies dealing with any aspect of allergenic molecule characterization and published in the last 40 years. The overall number of publications in this field is now around 6,600, and solely in the first 6 months of 2005 more than 500 papers have already been published on any aspect of allergens. Due to the increase of the number of identified allergens and their future use in clinical research, the number of published papers is expected to considerably increase in the near future. At the moment

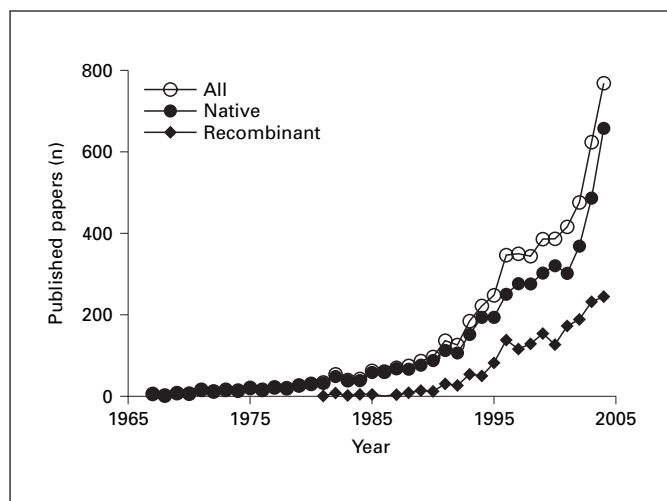


Fig. 3. Yearly published papers in international scientific journals dealing with at least one basic or clinical aspect of a defined allergen. Curves for papers dealing with native or recombinant allergens are given. Data are extracted from the Allergome database (www.allergome.org).

there are no specific tools, either literature databases or literature miners that are specific for the allergy field. The analysis of current web-based resources shows that only two of the web sites listed in table 1, InformAll and Allergome, report citations for each allergen to have an evidence-based documentation of reported data. This way of presenting citations is not or only partially dynamic and does not offer a cross-sectional view of literature data. For instance, the graph in figure 4 shows the unbalanced situation of experimental animal models using allergens. Studies using Gal d 2 (hen's egg ovalbumin), an allergen with little clinical relevance, account for 3 out of 4 papers published in the field. An evaluation of experimental animal models on the basis of the immunopathogenic mechanism to be explored leads to further unbalanced use of this molecule, whether Gal d 2 is a relevant allergen in humans or not (fig. 4). To produce these two graphs, references reported in each Allergome monographs had to be evaluated individually by keywording them on a separate reference database.

The rapidly expanding fields of allergen knowledge is prompting us to develop new computational methods for managing, analyzing and visualizing literature data [42, 43]. This approach should contribute to optimizing the exploitation of scientific data by providing investigators and clinicians with an easy interface with complex and large literature resources.

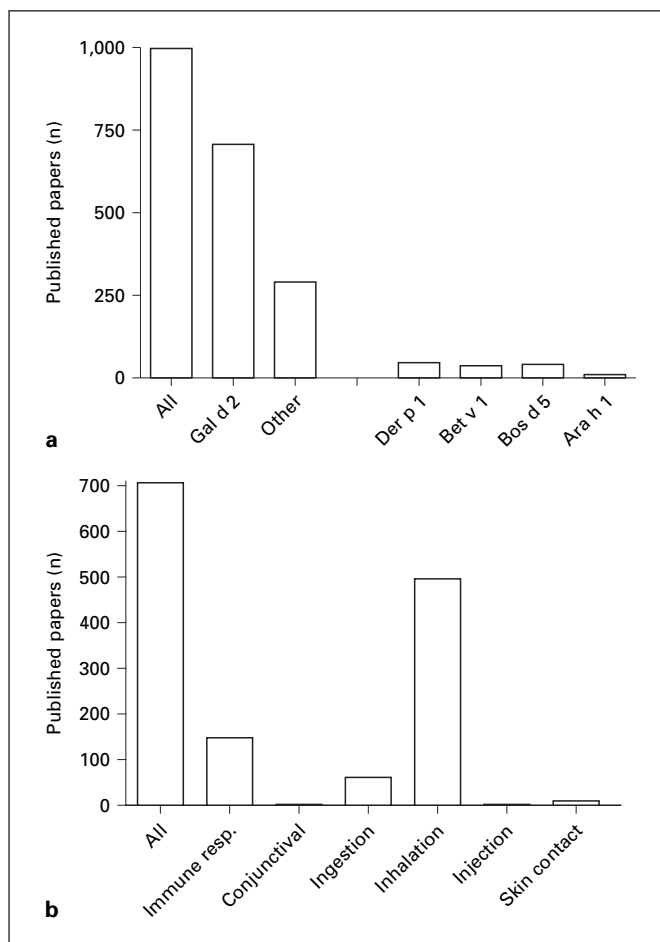


Fig. 4. **a** Published papers on experimental animal models considering most studied allergens in the last forty-years period. **b** Published papers on organ-specific experimental animal models considering Gal d 2 (hen's egg ovalbumin) in the last 40-year period. Data are extracted from the Allergome database (www.allergome.org).

Data Mining

There is a trend towards using the Internet for managing general medical information, and towards each individual possessing his or her electronic medical or health record. Some American companies (i.e. www.medem.com, www.medicallogic.com) are providing patient-centered web-based resources where personal health information and data are acquired and stored. Data can be retrieved by any authorized person who thus has a comprehensive overview of a patient's clinical history. This IT approach to medical data managing simplifies and speeds up the interactions between patient and physician and reduces health care costs for patients and healthcare

providers, and errors [44]. It is now evident that storage of this mass of information creates a mine of real time clinical data. For instance, the Medical Quality Improvement Consortium (<http://www.emr.msu.edu/>) allows members to access anonymous, patient de-identified, pooled data for research.

In the genomic field, various techniques have been developed that provide rapid, parallel surveys of gene expression patterns for hundreds or thousands of genes in a single assay. These transcriptional profiling techniques produce a wealth of data that can be used to develop a more complete understanding of gene function, regulation and interactions. Data mining of gene expression obtained from high-density microarray technology allows to store results from quantitative and simultaneous monitoring of thousands of genes by means of high-throughput assays, and generates a formidable mass of information. The data accumulated in the last 10 years are now stored [45] and analyzed by means of different approaches [46–48]. Allergies are the most common diseases worldwide and allergen-based diagnostic and epidemiological data are available from studies published in scientific journals after research projects have been set up and carried out. In an attempt to supply at least this information on allergenic molecules a static prevalence table is available in each Allergome monograph. Some prevalence data are reported for allergenic sources in the InformAll database in a text format. The use of IT in allergy data mining should be evaluated in the near future. Considering the source of data (skin test, IgE testing, provocation tests) great effort should be put in standardizing data source for data entry. Microarraying of highly purified allergenic molecules to be used in fully automated laboratory systems should lead us to real-time managing of allergy data. This could be the case very soon, at least for allergy diagnosis and epidemiology. A project to develop a web-based platform linked to the Allergome platform and called Real Time Monitoring of IgE sensitization (ReTiME) is going to be finalized in the forthcoming months. The ReTiME platform will allow uploading of raw diagnostic data from both in vivo and in vitro testing by any contributor worldwide. This should represent the first attempt to apply IT to allergy data mining.

Future Directions and Conclusions

Molecular biology, nanotechnology and IT offer a great opportunity to adapt and transform the use of currently available tools including pioneering web-based da-

tabases. There are two opposite possibilities: to either integrate allergy data within general-purpose databases or to create specific dynamic allergy databases spanning the entire knowledge in the field of the allergy. NCBI is a very well-known example of an integrated platform [49]. Through the global NCBI search engine (<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>) it is possible to retrieve information on the topic of interest at the same time from a platform of twenty-nine databases. It certainly is one of the most comprehensive free scientific resources on the web. As is the case for any general integrated platform, it requires more time to select relevant information that is anyway linked to other cross-related data, often from completely different biomedical fields. Much of the available information from this source can be downloaded for further independent processing. A further example of integration of data within the NCBI platform is the genomic and proteomic data mining database called GEO (Gene Expression Omnibus; <http://www.ncbi.nlm.nih.gov/geo/>) [45]. The GEO database facilitates effective mining and visualization of submitted genomic and proteomic data.

All these data are publicly available via the Internet, but none is specific to a biomedical branch.

Genomic analysis of allergenic sources generates a mass of data in the form of new allergenic structures to be classified [50–52]. An effort should be made to quickly integrate these data, having at the same time a list of unique nonredundant allergens, with isoforms and variants listed as part of their unique representative molecule [53]. This also represents a critical need for practical application of experimental findings.

Furthermore, a real integration should be reached with protein database to avoid overclassification as allergens of nonallergenic molecules. This is increasingly occurring when a novel protein structure identified from any organism is compared only for sequence homology to other proteins in these databases. This process, which often results in limited similarity leads to classifying as ‘potential’ allergens many structures without any clinical relevance. To overcome this trend, structure information should be more closely integrated with clinical data. Microarray-based experiments measuring IgE by means of purified native or recombinant molecules will add value in this direction. Combined information on a set of highly reliable allergen sequence should be available. These data sets should be publicly available.

Allergenicity should thus be evaluated using several tools [23] that can be obtained following the creation of powerful integrated databases storing sequences, biblio-

graphy, and raw data. Computational tools have to be applied to this dynamic mass of information to extract, interpret and visualize relevant new findings. Data will have to be curated by experts to define decision points. A critical aspect to be discussed is the business continuity of research groups responsible for allergen databases. As far as different users, such as clinicians, researchers, industries, authorities, institutions, are interested in maintaining a robust and comprehensive database platform, linked to already existing databases, combined support from many subjects should be considered.

In conclusion, as in many other biomedical fields, to derive most benefit from their work and further increase, organize, and manage the accumulated knowledge, scientists involved in the field of allergy should start using these specific and integrated new tools (fig. 2) for their research work in allergy and in clinical practice.

Notes and Acknowledgments

Data and info from Internet databases have been retrieved in July 2005. The author of this review is responsible for the Allergome project. I thank for their contribution to the Allergome project: Enrico Scala for scientific data processing, Gabriele Carabella and Daniele Riccioli for highly professional software developing, Anna Maria Ronconi and Virginia Mari for invaluable overall help in managing data. The Allergome platform is an independent project managed and financially supported by Allergy Data Laboratories s.c., a nonprofit company. The project is now also supported by unrestricted grants from Pharmacia Diagnostics (Uppsala, Sweden), Indoor Biotechnologies (Cardiff, UK), Laboratorios Leti (Barcelona, Spain), Diagnostic Product Corporation (Los Angeles, Calif., USA), Stallergenes (Antony, France), Allergopharma (Reinbek, Germany), VBC-Genomics (Vienna, Austria), and UCB Pharma (Pianezza, Italy). The Allergome project receives funding from the following institutions: University of Nebraska, 'Food Allergy Research and Resource Program' (Lincoln, Nebr., USA), University of Salzburg, Priority Programme 'BioScience and Health' (Salzburg, Austria), Institute for Research in Biomedicine (Bellinzona, Switzerland) and University 'Gabriele D'Annunzio' di Chieti e Pescara, Allergy and Clinical Immunology Section (Chieti, Italy).

References

- Kazemi-Shirazi L, Niederberger V, Linhart B, Lidholm J, Kraft D, Valenta R: Recombinant marker allergens: diagnostic gatekeepers for the treatment of allergy. *Int Arch Allergy Immunol* 2002;127:259–268.
- Chapman MD: Allergen nomenclature. *Clin Allergy Immunol* 2004;18:51–64.
- Breiteneder H, Radauer C: A classification of plant food allergens. *J Allergy Clin Immunol* 2004;113:821–830.
- Ferreira F, Hawranek T, Gruber P, Wopfner N, Mari A: Allergic cross-reactivity: from gene to the clinic. *Allergy* 2004;59:243–267.
- Harwanegg C, Laffer S, Hiller R, Mueller MW, Kraft D, Spitzauer S, Valenta R: Microarrayed recombinant allergens for diagnosis of allergy. *Clin Exp Allergy* 2003;33:7–13.
- Su X, Chew FT, Li SF: Piezoelectric quartz crystal based label-free analysis for allergy disease. *Biosens Bioelectron* 2000;15:629–639.
- Wiltshire S, O'Malley S, Lambert J, Kukanskis K, Edgar D, Kingsmore SF, Schweitzer B: Detection of multiple allergen-specific IgEs on microarrays by immunoassay with rolling circle amplification. *Clin Chem* 2000;46:1990–1993.
- Mullenix MC, Wiltshire S, Shao W, Kitos G, Schweitzer B: Allergen-specific IgE detection on microarrays using rolling circle amplification: correlation with in vitro assays for serum IgE. *Clin Chem* 2001;47:1926–1929.
- Kim TE, Park SW, Cho NY, Choi SY, Yong TS, Nahm BH, Lee S, Noh G: Quantitative measurement of serum allergen-specific IgE on protein chip. *Exp Mol Med* 2002;34:152–158.
- Bacarese-Hamilton T, Mezzasoma L, Ingham C, Ardizzoni A, Rossi R, Bistoni F, Crisanti A: Detection of allergen-specific IgE on microarrays by use of signal amplification techniques. *Clin Chem* 2002;48:1367–1370.
- Lebrun SJ, Petchpud WN, Hui A, McLaughlin CS: Development of a sensitive, colorimetric microarray assay for allergen-responsive human IgE. *J Immunol Methods* 2005;300:24–31.
- Hiller R, Laffer S, Harwanegg C, Huber M, Schmidt WM, Twardosz A, Barletta B, Becker WM, Blaser K, Breiteneder H, Chapman M, Cramer R, Duchene M, Ferreira F, Fiebig H, Hoffmann-Sommergruber K, King TP, Kleber-Janke T, Kurup VP, Lehrer SB, Lidholm J, Muller U, Pini C, Reese G, Scheiner O, Scheynius A, Shen HD, Spitzauer S, Suck R, Swoboda I, Thomas W, Tinghino R, Van Hage-Hamsten M, Virtanen T, Kraft D, Muller MW, Valenta R: Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J* 2002;16:414–416.
- Suck R, Nandy A, Weber B, Stock M, Fiebig H, Cromwell O: Rapid method for arrayed investigation of IgE-reactivity profiles using natural and recombinant allergens. *Allergy* 2002;57:821–824.
- Jahn-Schmid B, Harwanegg C, Hiller R, Bohle B, Ebner C, Scheiner O, Mueller M: Allergen microarray: comparison of microarray using recombinant allergens with conventional diagnostic methods to detect allergen-specific serum immunoglobulin E. *Clin Exp Allergy* 2003;33:1443–1449.
- Shreffler WG, Beyer K, Chu TH, Burks AW, Sampson HA: Microarray immunoassay: Association of clinical history, in vitro IgE function, and heterogeneity of allergenic peanut epitopes. *J Allergy Clin Immunol* 2004;113:776–782.
- Satinover SM, Reefer AJ, Pomes A, Chapman MD, Platts-Mills TA, Woodfolk JA: Specific IgE and IgG antibody-binding patterns to recombinant cockroach allergens. *J Allergy Clin Immunol* 2005;115:803–809.
- Galperin MY: The Molecular Biology Database Collection: 2005 update. *Nucleic Acids Res* 2005;33:D5–D24.
- Gendel SM: Sequence databases for assessing the potential allergenicity of proteins used in transgenic foods. *Adv Food Nutr Res* 1998;42:63–92.
- Ivanciu O, Schein CH, Braun W: SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 2003;31:359–362.
- Mari A, Riccioli D: The Allergome Web Site – A Database of Allergenic Molecules. Aim, Structure, and Data of a web-based resource. *J Allergy Clin Immunol* 2004;113:S301.
- Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP, Van Ham RC: Allermatch TM, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* 2004;5:133.
- Riaz T, Hor HL, Krishnan A, Tang F, Li KB: WebAllergen: a web server for predicting allergenic proteins. *Bioinformatics* 2005;21:2570–2571.

- 23 Spok A, Gaugitsch H, Laffer S, Pauli G, Saito H, Sampson H, Sibanda E, Thomas W, Van Hage M, Valenta R: Suggestions for the Assessment of the Allergenic Potential of Genetically Modified Organisms. *Int Arch Allergy Immunol* 2005;137:167–180.
- 24 Goodman RE, Hefle SL, Taylor SL, van Ree R: Assessing Genetically Modified Crops to Minimize the Risk of Increased Food Allergy: A Review. *Int Arch Allergy Immunol* 2005;137:153–166.
- 25 Bublin M, Mari A, Ebner C, Knulst A, Scheiner O, Hoffmann-Sommergruber K, Breiteneder H, Radauer C: IgE sensitization profiles toward green and gold kiwifruits differ among patients allergic to kiwifruit from 3 European countries. *J Allergy Clin Immunol* 2004;114:1169–1175.
- 26 Prescott RA, Potter PC: Hypersensitivity to airborne spitting cobra snake venom. *Ann Allergy Asthma Immunol* 2005;94:600–603.
- 27 Binder M, Mahler V, Hayek B, Sperr WR, Scholler M, Prozell S, Wiedermann G, Valent P, Valenta R, Duchene M: Molecular and immunological characterization of arginine kinase from the Indian meal moth, *Plodia interpunctella*, a novel cross-reactive invertebrate pan-allergen. *J Immunol* 2001;167:5470–5477.
- 28 Bolhaar ST, van Ree R, Bruijnzeel-Koomen CA, Knulst AC, Zuidmeer L: Allergy to jackfruit: a novel example of Bet v 1-related food allergy. *Allergy* 2004;59:1187–1192.
- 29 Mothes N, Horak F, Valenta R: Transition from a botanical to a molecular classification in tree pollen allergy: implications for diagnosis and therapy. *Int Arch Allergy Immunol* 2004;135:357–373.
- 30 Larsen JN, Lowenstein H: Allergen nomenclature. *J Allergy Clin Immunol* 1996;97:577–578.
- 31 Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TA: Allergen nomenclature. *Bull World Health Organ* 1986;64:767–774.
- 32 King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W: Allergen Nomenclature. *J Allergy Clin Immunol* 1995;96:5–14.
- 33 Zola H, Swart B, Nicholson I, Aasted B, Bensusan A, Boumsell L, Buckley C, Clark G, Drbal K, Engel P, Hart D, Horejsi V, Isacke C, Macardle P, Malavasi F, Mason D, Olive D, Saalmueller A, Schlossman SF, Schwartz-Albiez R, Simmons P, Tedder TF, Uguccioni M, Warren H: CD molecules 2005: human cell differentiation molecules. *Blood* 2005;DOI 10.1182/blood-2005-03-1338.
- 34 Ivanciuc O, Schein CH, Braun W: Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics* 2002;18:1358–1364.
- 35 Ivanciuc O, Mathura V, Midoro-Horiuti T, Braun W, Goldblum RM, Schein CH: Detecting potential IgE-reactive sites on food proteins using a sequence and structure database, SDAP-Food. *J Agric Food Chem* 2003;51:4830–4837.
- 36 Hileman RE, Silvanovich A, Goodman RE, Rice EA, Holleschak G, Astwood JD, Hefle SL: Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int Arch Allergy Immunol* 2002;128:280–291.
- 37 Stadler MB, Stadler BM: Allergenicity prediction by protein sequence. *FASEB J* 2003;17:1141–1143.
- 38 Mari A, Truffer R, Stadler MB, Stadler BM: Identification of allergenic molecules using the 'Sequence Motifs' Computational Approach. *Allergy Clin Immunol Int* 2005;17:30.
- 39 Mari A, Scala E, Ronconi AM, Carabella G: Sequence homology and IgE co-recognition of allergenic molecules: comparative results obtained using the Allergome O-Ring Graphical Representation. *Allergy Clin Immunol Int* 2005;17:205.
- 40 Bjorklund AK, Soeria-Atmadja D, Zorzet A, Hammerling U, Gustafsson MG: Supervised identification of allergen-representative peptides for in silico detection of potentially allergenic proteins. *Bioinformatics* 2005;21:39–50.
- 41 Haynes RB, McKibbin KA, Wilczynski NL, Walter SD, Werre SR: Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ* 2005;330:1179.
- 42 Chaussabel D: Biomedical literature mining: challenges and solutions in the 'omics' era. *Am J Pharmacogenomics* 2004;4:383–393.
- 43 Chaussabel D, Sher A: Mining microarray expression data by literature profiling. *Genome Biol* 2002;3:Research0055.
- 44 Oliven A, Michalake I, Zalman D, Dorman E, Yeshurun D, Odeh M: Prevention of prescription errors by computerized, on-line surveillance of drug order entry. *Int J Med Inform* 2005;74:377–386.
- 45 Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R: NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res* 2005;33:D562–D566.
- 46 Schulze A, Downward J: Analysis of gene expression by microarrays: cell biologist's gold mine or minefield? *J Cell Sci* 2000;113 pt 23:4151–4156.
- 47 Quackenbush J: Computational analysis of microarray data. *Nat Rev Genet* 2001;2:418–427.
- 48 Schulze A, Downward J: Navigating gene expression using microarrays – a technology review. *Nat Cell Biol* 2001;3:E190–E195.
- 49 Pruitt KD, Tatusova T, Maglott DR: NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2005;33:D501–D504.
- 50 Angus AC, Ong ST, Chew FT: Sequence tag catalogs of dust mite-expressed genomes: utility in allergen and acarologic studies. *Am J Pharmacogenomics* 2004;4:357–369.
- 51 Hong SG, Cramer RA, Lawrence CB, Pryor BM: Alt a 1 allergen homologs from *Alternaria* and related taxa: analysis of phylogenetic content and secondary structure. *Fungal Genet Biol* 2005;42:119–129.
- 52 Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer I, van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ: Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet* 2005;111:171–183.
- 53 Brusica V, Millot M, Petrovsky N, Gendel SM, Gigonzac O, Stelman SJ: Allergen databases. *Allergy* 2003;58:1093–1100.